



Munich Personal RePEc Archive

Testing for Separation in Agricultural Household Models and Unobservable Household-Specific Effects

Jean-Louis Arcand and Beatrice d'Hombres

2006

Online at <http://mpa.ub.uni-muenchen.de/1863/>
MPRA Paper No. 1863, posted 21. February 2007

Testing for Separation in Agricultural Household Models and Unobservable Household-Specific Effects

August 28, 2006

Abstract

When market structure is complete, factor demands by households will be independent of their characteristics, and households will take their production decisions as if they were profit-maximizing firms. This observation constitutes the basis for one of the most popular empirical tests for complete markets, commonly known as the “separation” hypothesis. In this paper, we show that all existing tests for separation using panel data are potentially biased towards rejecting the null-hypothesis of complete markets, because of the failure to adequately control for unobservable individual effects. Since the variable on which the test for separation is based cannot be identified in most panel datasets following the usual covariance transformations, and is likely to be correlated with the household-specific effect, neither the within nor the variance-components procedures are able to solve the problem. We show that the Hausman-Taylor 1981 estimator, in which the impact of covariates that are invariant along one dimension of a panel can be identified through the use of covariance transformations of other included variables that are orthogonal to the household-specific effects as instruments, provides a simple solution. Our approach is applied to a rich Tunisian dataset in which separation -and thus the null of complete markets- is strongly rejected using the standard approach, but is not rejected once correlated unobservable household-specific effects are controlled for using the Hausman-Taylor instrument set.

Keywords: panel data, household-specific effects, household models, testing for incomplete markets, development microeconomics, Tunisia.

JEL: O120, C230, D130, D520.

One of the most widely-used empirical tests for the presence of market imperfections in developing countries is provided by the so-called “separation” hypothesis. Numerous papers, including the seminal article by Benjamin (1992), have tested the hypothesis that factor demands on a given plot of land will be independent of household characteristics, when market structure is (almost) complete. The early literature is well summarized in Singh, Squire, and Strauss (1986), while Udry (1996a) covers the more recent literature as well as providing two careful applications to African data.

Separation implies that the marginal productivity of inputs will be a function solely of plot characteristics and prices, and that households take their production decisions as if they were profit-maximizing firms. In contrast, when factor demands are a function of household characteristics, marginal productivities are not equated across households and production is inefficient.

The purpose of this article is to demonstrate that: (i) in most cases, the standard test for separation using panel data is biased towards rejecting the null-hypothesis of complete markets because of a problem of unobservable household-specific effects; (ii) the usual covariance transformations performed on panel data cannot solve this problem; but (iii) the Hausman and Taylor (1981) estimator can. We show, using a rich plot-level Tunisian dataset, that the null-hypothesis of complete markets is rejected using the standard approach, while it is not once correlated household-specific, time-varying effects are controlled for using the Hausman-Taylor estimator.

In a plot-level version of the test for separation, the equation being estimated on agronomic data is given by:

$$(1) \quad Y_{iht} = X_{iht}\alpha + Z_{ht}\beta + \varepsilon_{iht},$$

where Y_{iht} is total labor usage (i.e., family and hired labor) on plot i , cultivated

by household h , at time t , X_{iht} is a matrix of plot characteristics, Z_{ht} is a matrix of household characteristics, and ε_{iht} is a disturbance term that satisfies the usual Gauss-Markov assumptions. Separation is then associated with a simple F -test on the exclusion restriction that $\beta = 0$. In the common expression of the test for separation (Benjamin 1992), Z_{ht} is household size.

The main problem associated with this procedure is that the disturbance term ε_{iht} can be decomposed into a nested error component structure given by:

$$(2) \quad \varepsilon_{iht} = \mu_t + \lambda_h + \lambda_{ht} + \eta_{iht},$$

where μ_t is a shock common to all plots and households at time t , λ_h is a time-invariant household effect, λ_{ht} is a household-time effect, and η_{iht} is a disturbance term that satisfies the usual assumptions (see Baltagi, Song and Jung 2001).¹ In most plot-level datasets used in the literature, each household cultivates several plots. This is a standard panel data framework, with one dimension being given by plots, the second by households, and the third by time. Although λ_h can be accounted for by a “within” procedure which transforms variables into deviations with respect to their household-specific means (over *all* time periods), there remains λ_{ht} . Since it is probable that λ_{ht} is correlated with Z_{ht} , the least-squares estimate of β , even after the standard “within” transformation, will be biased with, in the scalar case:

$$(3) \quad \text{p lim } \hat{\beta}_w = \beta + \frac{\text{cov}[\lambda_{ht}, \hat{\varepsilon}_{iht}]}{\sigma_e^2},$$

where σ_e^2 is the variance of the residual $\hat{\varepsilon}_{iht}$ from the auxiliary “within” regression of household size on X_{iht} (see Hsiao 1986, p. 64, equation (3.9.3)).² If $\text{cov}[\lambda_{ht}, \hat{\varepsilon}_{iht}] \neq 0$, as is likely in the context of what is essentially a labor demand equation, then *all standard tests of separation are biased towards rejecting the null-hypothesis of complete markets*, when the “true” value of β is zero. One may therefore reject the null not

because market structure is necessarily incomplete, but simply because of a banal problem of unobservable heterogeneity. Another way of putting this is that, in the standard test, the rejection of separation is conditional on the maintained identifying assumption that λ_{ht} is the same across all households at a given time t . It is very likely that this assumption is violated.

The usual econometric response to a problem of unobservable individual heterogeneity in panel data is to apply one of the standard covariance transformations, such as the “within” procedure. Here, this would involve expressing all variables as deviations with respect to their household-specific means, *at a given t* . While, under the assumption of exogeneity, this does allow one to recover unbiased estimates of α , it has the regrettable side-effect of eliminating the variable(s) upon which the test for separation is based since, when one sweeps out λ_{ht} , one also sweeps out Z_{ht} . Since it is highly likely that λ_{ht} is not orthogonal to Z_{ht} , random effects are not an answer, as they too will yield biased estimates of β .

Moreover, standard instrumental variables procedures, in which one would simply instrument for Z_{ht} , are not usually implementable. This is because admissible exogenous instruments that would be correlated with Z_{ht} but are orthogonal to λ_{ht} are usually not available or, if they are, should probably already be included in Z_{ht} for theoretical reasons.

The problem, which is similar in spirit to that of consistently estimating the returns to education using panel data when schooling is correlated with the individual effects, can be solved using the Hausman-Taylor (1981, henceforth, HT) instrumental variables estimator, which allows one to control for unobservable individual effects that are correlated with Z_{ht} , while allowing one to identify β .

A simple household model

The Pareto-optimal baseline

The following model is in part inspired by Udry (1996b). A useful primer on household models is provided by Bardhan and Udry (1999). There are two types of family members, men and women, and two types of hired labor, male and female. This corresponds to the situation in the Tunisian village that will be the object of our empirical analysis.

Consider a household, indexed by h , constituted by two members indexed by $j = M, F$, that cultivates several plots of land, indexed by $i = 1, \dots, I_h$. Individual j consumes a quantity c_k^j of good $k = 1, \dots, K$; c^j is therefore the $1 \times K$ vector of private goods consumed by individual j , whereas total household consumption is given by the $1 \times K$ vector $c = c^M + c^F$. Total labor supply of individual j is equal to L^j . Public goods produced within the household are given by Z . Preferences of individual j are given by: $U^j = U^j(c^M, c^F, Z, L^M, L^F, \Omega)$, where Ω is a vector of household taste shifters such as household demographics or wealth. Consider a plot, indexed by i , that is cultivated in crop k . Then output on such a plot is given by $q_i^k = F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)$, where A_i represents the characteristics of plot i , including plot size, soil type and irrigation status, and H_i^j (L_i^j) is hired (family) labor of sex j used on the plot. Using the notation in Udry (1996b), where the set of plots cultivated in crop k are denoted by $P^k = \{i | \text{crop } k \text{ is grown on plot } i\}$, the total production of crop k by the household is given by:

$$(4) \quad q^k = \sum_{i \in P^k} q_i^k = \sum_{i \in P^k} F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i), k = 1, \dots, K,$$

and the $1 \times K$ vector of outputs of all goods is given by $q = (q^1, q^2, \dots, q^K)$.

Public good production within the household is given by:

$$(5) \quad Z = Z(L_Z^M, L_Z^F).$$

The time constraint of household member j is given by:

$$(6) \quad L^j = L_Z^j + \sum_i L_i^j + L_W^j, j = M, F,$$

where L_W^j is household member's j labor supply on the labor market. Finally, letting $p = (p^k, \dots, p^K)$ denote the $1 \times K$ vector of prices, the household's budget constraint is given by:

$$(7) \quad pc' \leq pq' - w^M \sum_i H_i^M - w^F \sum_i H_i^F + w^M L_W^M + w^F L_W^F + I,$$

where w^j denotes the wage rate paid to hired labor of sex j , and I is non-labor income. From the usual corollary to the First Theorem of Welfare Economics (see e.g. Varian 1978), and for any Pareto weight $\lambda > 0$, the intra-household allocation of resources will be Pareto-optimal as long as it solves the problem:

$$(8) \quad \max_{\{c, L, H, P^k\}} U^M + \lambda U^F \quad s.t. \quad (4), (5), (6) \text{ and } (7).$$

Substituting the binding constraints ((4),(5) and (6)) into the objective function, and letting μ denote the Lagrange multiplier associated with (7), the necessary first-order conditions (FOCs) are then given by:

$$(9) \quad \frac{\partial U^M}{\partial c_k^j} + \lambda \frac{\partial U^F}{\partial c_k^j} - \mu p_k = 0,$$

for $k = 1, \dots, K, j = M, F$;

$$(10a) \quad \left(\frac{\partial U^M}{\partial Z} + \lambda \frac{\partial U^F}{\partial Z} \right) \frac{\partial Z(L_Z^M, L_Z^F)}{\partial L_Z^j} + \frac{\partial U^M}{\partial L^j} + \lambda \frac{\partial U^F}{\partial L^j} = 0,$$

$$(10b) \quad \frac{\partial U^M}{\partial L^j} + \lambda \frac{\partial U^F}{\partial L^j} + \mu w^j = 0,$$

for $j = M, F$;

$$(11a) \quad \frac{\partial U^M}{\partial L^j} + \lambda \frac{\partial U^F}{\partial L^j} + \mu p_k \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = 0,$$

$$(11b) \quad \mu p_k \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} - \mu w^j = 0,$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$, plus the complementary slackness condition from the Kuhn-Tucker Theorem.

Combining (10b), (11a) and (11b) then implies that the marginal productivity of family labor of sex j will be equated to the marginal productivity of hired labor of the same sex, and that these marginal productivities will be the same across all plots cultivated by a given household, as well as *between* households:

$$(12) \quad \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} = \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = \frac{w^j}{p_k},$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$.

Market imperfections

The literature on household models is replete with examples of market imperfections that lead to violations of the conditions that underly separability. In what follows, we consider the most commonly appealed to market imperfections and examine their consequences on the optimality conditions derived above. These include credit con-

straints (Eswaran and Kotwal 1986; Feder 1985; Feder et al. 1990), labor market imperfections (Lopez 1984; Benjamin 1992; Jacoby 1993; Skoufias 1994; Lambert and Magnac 1994; Sadoulet, DeJanvry and Benjamin 1998; Sonoda and Maruyama 1999; Bowlus and Sicular 2003), imperfect land rental markets or tenure rights (Gavian and Fafchamps 1996; Carter and Yao 2002; Carter and Olinto 2003 on the interaction between land and labor market imperfections), imperfect insurance markets in conjunction with imperfect labor and land markets (Barrett 1996; Kevane 1996), or marketing constraints.

As noted by Udry (1996a), the separation result is robust to the absence of the market for one of the factor inputs.

Credit constraints

Consider now the model given in (8), to which we append a working capital constraint of the form $w^M \sum_i H_i^M + w^F \sum_i H_i^F - w^M L_W^M - w^F L_W^F - I \leq B$, where B is the working capital available to the household. Denoting the Lagrange multiplier associated with the credit constraint by φ , the ensuing FOCs imply that:

$$(13) \quad \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = \frac{\partial F^k(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} = \left(\frac{\mu + \varphi}{\mu} \right) \frac{w^j}{p_k},$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$. This condition implies that the marginal productivities of family and hired labor of sex j are equated across plots cultivated by a given household. In contrast to the separable case, on the other hand, these marginal productivities differ *between* households, because of the presence of the Lagrange multiplier φ , which is household-specific.

Labor market imperfections

Consider now a constraint on the amount of labor that a household can "export" on the labor market: $L_W^j \leq \bar{L}_W^j$. Then, denoting the Lagrange multipliers associated with these constraints by ψ^j , the ensuing FOCs imply that:

$$(14a) \quad \frac{\partial F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} = \frac{w^j}{p_k},$$

$$(14b) \quad \frac{\partial F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = \frac{w^j}{p_k} - \frac{\psi^j}{\mu p_k},$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$. The first condition implies that the marginal productivity of hired labor will be equated across plots cultivated by the same household, as well as across plots cultivated by different households. The second condition implies that the marginal productivity of family labor will be equated across plots cultivated by a given household, but will not be equated across households. Moreover, the marginal productivities of family and hired labor will not be equated within households.

Now consider a constraint on the other side of labor market that takes the form of a limit on the amount of labor that the household can hire: $\sum_i H_i^j \leq \bar{H}^j$. Then, denoting the Lagrange multiplier associated with these constraints by ψ^j , the FOCs associated with the problem imply that:

$$(15a) \quad \frac{\partial F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} = \frac{w^j}{p_k} + \frac{\psi^j}{\mu p_k},$$

$$(15b) \quad \frac{\partial F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = \frac{w^j}{p_k},$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$. These conditions are the mirror image of those given in the case of labor exports. The first condition implies that the marginal productivity of hired labor will be equated across plots cultivated by a

given household, but will differ between households. The second condition implies that the marginal productivity of family labor will be equated across plots cultivated by a given household, as well as between households. As with the constraint on the labor export side, the marginal productivities of family and hired labor will not be equated within households.

Marketing constraints

Consider now a constraint that takes the form of an upper bound \overline{Q}^l on the amount of crop l that the household can sell. More formally, the constraint in question can be written as $\sum_{i \in P^l} F^l(L_i^M, L_i^F, H_i^M, H_i^F, A_i) - (c_i^M + c_i^F) \leq \overline{Q}^l$. Letting ϕ^l denote the Lagrange multiplier associated with the constraint, the FOCs that correspond to the problem then imply that, for those plots on which crop l is grown:

$$(16) \quad \frac{\partial F^l(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} = \frac{\partial F^l(L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} = \frac{\mu w^j}{\mu p_k - \phi^l},$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$. These conditions imply that the marginal productivities of family and hired labor are equated across plots cultivated by a given household in crop l , but that these marginal productivities will differ between households. For other crops $k \neq l$ that are not subject to the marketing constraint, the conditions given in the unconstrained case continue to hold.

Insurance market failure

Assume now that the production technology on plot i is given by:

$$(17) \quad q_i^k = F^k(\theta_i, L_i^M, L_i^F, H_i^M, H_i^F, A_i),$$

where θ_i is a stochastic shock to production on plot i . If we denote the vector of stochastic shocks affecting all of the plots cultivated by household h by $\theta_h = (\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_{I_h})$ which is assumed to be distributed according to the joint probability density function (*pdf*) $g(\theta)$ then the household's optimization problem is given by:

$$(18) \quad \max_{\{c, L, H, P^k\}} E_{\theta_h} [U^M + \lambda U^F] \quad s.t. \quad (4), (5), (6) \text{ and } (7),$$

where $E_{\theta_h} [U^M + \lambda U^F] = \int \dots \int (U^M + \lambda U^F) g(\theta) d\theta_1 d\theta_2 \dots d\theta_i \dots d\theta_{I_h}$. In the absence of an insurance market that would allow the household to equate the marginal utility of its consumption across states of nature, the associated FOCs which implicitly define optimal labor inputs will then be given by:

$$(19a) \quad E_{\theta_h} \left[\left(\frac{\partial U^M}{\partial c_1^M} + \lambda \frac{\partial U^F}{\partial c_1^M} \right) \left(\frac{\partial F^k(\theta_i, L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial L_i^j} - \frac{w^j}{p_k} \right) \right] = 0,$$

$$(19b) \quad E_{\theta_h} \left[\left(\frac{\partial U^M}{\partial c_1^M} + \lambda \frac{\partial U^F}{\partial c_1^M} \right) \left(\frac{\partial F^k(\theta_i, L_i^M, L_i^F, H_i^M, H_i^F, A_i)}{\partial H_i^j} - \frac{w^j}{p_k} \right) \right] = 0,$$

for $k = 1, \dots, K, j = M, F, i = 1, \dots, I_h$. The upshot is that insurance market failure implies that optimal input use on plot i is a function not only of plot i 's characteristics but, through the marginal utility of consumption $\frac{\partial U^M}{\partial c_1^M} + \lambda \frac{\partial U^F}{\partial c_1^M}$, of the characteristics of *all* of the plots cultivated by the household.

The tenancy market

Consider now a situation, as is the case in the Tunisian village that will be the focus of the empirical portion of this paper, in which there is an active land rental market in which sharecropping and fixed rental contracts arise. Let

$$(20) \quad P^{km} = \{i | \text{crop } k \text{ is grown on plot } i \text{ under a contract of type } m\},$$

where $m = OO$ (owner operator), RI (rented in), SI (sharecropped in), SO (sharecropped out), and

$$(21) \quad P^n = \{i \mid \text{plot } i \text{ is rented under a contract of type } n\},$$

where $n = RI$ (rented in), RO (rented out). When a household rents in a plot i under a sharecropping contract it retains a fraction α_i of output and pays a fraction β_i of the costs associated with the plot; when it rents a plot in under a fixed rental contract, it is residual claimant and *pays* a fixed rental equal to R_i ; when it rents out a plot under a sharecropping contract, it retains a fraction $1 - \alpha_i$ of output and pays a fraction $1 - \beta_i$ of costs; finally, when a household rents out a plot under a fixed rental contract, it *receives* a fixed rental payment equal to R_i . The household's budget constraint is therefore given by:

$$(22) \quad \sum_k p_k (c_k^M + c_k^F) \leq \sum_k \left(\sum_{i \in P^{kOO} \cup P^{kRI}} \begin{bmatrix} p_k F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i) \\ -w^M H_i^M - w^F H_i^F \end{bmatrix} \right. \\ \left. + \sum_{i \in P^{kSI}} \begin{bmatrix} p_k \alpha_i F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i) \\ -\beta_i (w^M H_i^M - w^F H_i^F) \end{bmatrix} \right. \\ \left. + \sum_{i \in P^{kSO}} \begin{bmatrix} p_k (1 - \alpha_i) F^k (L_i^M, L_i^F, H_i^M, H_i^F, A_i) \\ -(1 - \beta_i) (w^M H_i^M - w^F H_i^F) \end{bmatrix} \right) \\ - \sum_{i \in P^{RI}} R_i + \sum_{i \in P^{RO}} R_i + w^M L_W^M + w^F L_W^F + I.$$

Note that the household chooses factor input use on those plots (i) that it cultivates as an owner-operator and (ii) that it rents in either under a fixed rental or a sharecropping contract.

Within-household inefficiency

Empirical implementation

The Hausman-Taylor instrument set

Let X_{1iht} be those elements of X_{iht} that are uncorrelated with λ_{ht} , while X_{2iht} are those that are; Z_{1ht} and Z_{2ht} are defined in a similar manner. The set of instruments proposed by HT (1981), adapted to the three-dimensional panel structure, is $A_{HT} = [Q_{vt}X_{iht}; P_{vt}X_{1iht}; Z_{1ht}]$, where P_{vt} and Q_{vt} are the idempotent matrices that perform the “between” and “within” transformations at time t , respectively.³ Under the assumption that X_{iht} is uncorrelated with η_{iht} , $Q_{vt}X_{iht}$ is a legitimate set of instruments since $E[(Q_{vt}X_{iht})'\eta_{iht}] = 0$. The basic intuition behind the HT estimator is that only the λ_{ht} component of the error term is correlated with $[X_{2iht} \ Z_{2ht}]$, which allows one to use $Q_{vt}X_{2iht}$ as instruments for X_{2iht} , while $P_{vt}X_{1iht}$ furnishes the instruments for Z_{2ht} . The HT estimator therefore allows one to control for unobservable correlated individual effects, while allowing one to identify the parameters of interest (β) in the context of testing for separation. A necessary condition for identification is that the number of elements of X_{1iht} be greater than the number of elements of Z_{2ht} (HT 1981, PROPOSITION 3.2, p. 1385). These results have been extended by Amemya and MaCurdy (1986) and Breusch, Mizon, and Schmidt (1989) who suggest a broader set of instruments that should improve efficiency. Their approach, however, is only possible on balanced data, which is not the case in the dataset used in this paper or, for that matter, in most plot-level agronomic datasets. Notice that the HT instrument set is admissible only if exogeneity is satisfied. This is another potential source of bias in tests for separation, but which is difficult to address because of the lack of admissible plot-level instruments in most datasets.

The three-dimensional nature of our dataset allows us an additional degree of

freedom in terms of the definition of HT-type instruments. Above, we considered orthogonality conditions of subsets of X_{iht} and Z_{ht} with respect to the individual-time effect λ_{ht} . But the three-dimensional nature of the data also allows us to construct instruments based on orthogonality conditions with respect to variables that have been purged of their time-invariant household-specific component which is correlated with λ_h . An advantage of this procedure is that, in empirical applications, the orthogonality of $P_{vt}X_{1iht}$ with respect to the individual effects could be suspect. Purging $P_{vt}X_{1iht}$ of its component that is correlated with the time-invariant individual effect, λ_h , should render it more palatable as a potential instrument set. In that case, the set of HT-type instruments is given by $\left[Q_{vt}X_{1iht}; Q_{vt}X_{2iht}; \tilde{X}_{1iht}; Z_{1ht}\right]$, where $\tilde{X}_{1iht} = Q_v(P_{vt}X_{1iht})$ denotes the matrix of explanatory variables that have been purged of their component which is correlated with λ_h . In other words, $Q_v = I - P_v$ is the idempotent projection matrix that transforms variables into deviations with respect to their household-specific means (over all time periods).⁴

We instrument contractual choice as .

Household characteristics

Reduced form estimates

The basic estimating equation is given by:

$$(23) \quad X_{ih}^j = A_{ih}\delta + \Omega_h\gamma + \eta_h + \varepsilon_{ih}, X = L, H, j = M, F.$$

plot characteristics are given by $A_{ih}\delta = \sum_{m=1}^{m=M} \left(\frac{\beta_{mih}}{\alpha_{mih}} \right) \delta_m + \sum_s S_{sih}\delta_s + \sum_k d_k\delta_k + T_{ih}\delta_T$

HausmanTaylor estimation cited by referee: Baltagi and Khanti-Akom (1990), Cornwell and Rupert (1988).

Estimation of productivity differentials using panel data cited by referee: Deininger and Olinto (2000), Gardebreek and Lansink (2003).

In order to illustrate our fundamental point concerning the bias affecting conventional tests for separation in household models, consider the following standard procedure implemented on a typical plot-level dataset. The data come from two surveys (1993, 1995) carried out in the village of El Oulja, Tunisia (see Matoussi and Nugent 1989, and Laffont and Matoussi 1995, for descriptions of the village). These data display those properties discussed in the introduction: a Hausman test of random household-time effects (λ_{ht}) versus fixed effects in an empirical counterpart to equations (1) and (2) strongly rejects (with a p -value below 0.001) the null of the absence of correlation between λ_{ht} and Z_{ht} . The bias identified in equation (3) is therefore manifestly present in conventional tests of the null-hypothesis of complete markets using this panel dataset.

For the purpose of HT estimation, we divide the explanatory variables into two categories: (i) X_{1iht} variables, assumed to be uncorrelated with λ_{ht} , include four soil type dummies and a dummy variable that indicates whether the plot is irrigated or not, as well as a set of eight crop dummies;⁵ (ii) X_{2iht} variables, assumed to be correlated with λ_{ht} , are given by the share of costs borne by the cultivator, divided by the share of output received, for eight different inputs, as well as log plot size in hectares.⁶

The economic rationale for allowing the variables included in X_{2iht} to be correlated with λ_{ht} is that they may, in the context of tenancy contracts (which account for 28 percent of the plots in the sample), be determined as the solution to a principal-agent relationship between a landlord and a tenant, and would then be functions of tenant characteristics, including those unobservable characteristics potentially captured by λ_{ht} . An additional, empirical, motivation for using the ratios of cost-shares to the output share is that the data in question come from a single village and that the only source of cross-sectional variation in effective input prices stems from heterogeneity in contractual form on plots under tenancy contracts. Plot size is also assumed to

be correlated with λ_{ht} , as it too may be chosen by landlords for plots under tenancy contracts. Both of these hypotheses will be subjected to a test of the corresponding overidentifying restrictions below. Our single Z_{2ht} variable is given by log household size. In line with the usual methodology, the dependent variable is log total (hired and family) labor usage on the plot, in person-days per hectare.⁷ Table 1 provides summary statistics on all the aforementioned variables.

Estimation results are presented in table 2. Many households did not engage in crop production in the second survey (1995) because of adverse climatic shocks; this explains why the number of household-years (ht) is much smaller than *twice* the number of households (h). Intercept, year dummy, and eight crop dummies are included in all specifications presented in table 4 (no constant in col. (2)); random effects are rejected in favor of fixed effects in columns (2) and (3) by Hausman tests with associated p-values of less than 0.0001. The dependent variable in the estimation results presented in table 4 is given by log person-days per hectare used on the plot.

The standard test for separation is presented in column 1, and yields an unambiguous rejection of the null-hypothesis of complete markets in that log household size is highly significant at the usual levels of confidence (t -statistic = 4.55). In column 2, we control for time-invariant household characteristics (λ_h) using the “within” transformation: recall that the impact of household size can be identified here because we have two years of data and household size varies over the two surveys. Note, despite a substantial fall in the variance of log household size, which goes from 0.328 in levels, to 0.014 when expressed in terms of deviations with respect to household-specific means (over both periods), that the estimated standard error is still reasonably small, with the associated t -statistic being equal to 2.406. The time-invariant household fixed effects (λ_h) used here corresponds to the type of specification used by Udry (1996b), table 3, column 2, for a labor demand per hectare equation estimated on the Burkina Faso ICRISAT dataset. Again the null of complete markets is strongly rejected by the

data (t -statistic = 2.56).⁸ In column 3, we present results which allow for random household-time (λ_{ht}) effects: this specification, which also rejects the null of complete markets, can however be dismissed on the basis of the corresponding Hausman test in favor of fixed effects, as mentioned above (p -value of the Hausman test is below 0.001). Of course, household-time (λ_{ht}) *fixed* effects would not allow one to test for separation at all in that they would also sweep out the impact of household size.

In column 4, we present results corresponding to the efficient HT estimator.⁹ The results are striking. In contrast to what was found in columns 1 through 3, the null of complete markets is *not* rejected at the usual levels of confidence: the point estimate of the parameter associated with household size is statistically indistinguishable from zero (t -statistic = 0.71).¹⁰ Moreover, the test of the overidentifying restrictions does not lead one to reject, with a p -value equal to 0.545. In addition, the Shea partial R^2 (0.071 for the reduced form for household size) and the partial F -test (p -value below 0.001) of the joint significance of the instruments indicate that we are not facing a "weak instruments" problem.¹¹ We also compute the test (Bowden and Turkington 1984) based on canonical correlation, which rejects the null that the smallest canonical correlation is zero (p -value = 0.002). Hall, Rudebusch, and Wilcox (1996) show that IV estimators are not weakly identified if and only if all the canonical correlations converge to non-zero limits. They develop a likelihood ratio statistic for the null that the smallest canonical correlation is zero.

However, the Hansen test is potentially inconclusive insofar as this test is based on the strong assumption that at least as many instruments as the number of elements of Z_{2ht} are exogenous. As the Hausman-Taylor procedure is very sensitive to the choice of the variables included in X_{1iht} and X_{2iht} , we compute the "difference Hansen" statistic which enables us to test the validity of subsets of instruments (Hayashi 2000). To that end, we first implement the HT estimator using \tilde{X}_{1iht} in addition to the basic matrix of instruments used earlier. Second, we test that the subset of instruments

$P_{vt}X_{iht}$ satisfies the orthogonality conditions. The "difference Hansen" statistic presented in table 2 does not lead one to reject the null hypothesis that the specified variables are admissible instruments (p -value = 0.676).¹²

The upshot is that, in stark contrast to the usual approach which does not control for unobservable individual effects, HT estimation leads to the non-rejection of the null-hypothesis of complete markets. Moreover the consistency of the HT-based results presented in column 4 is ensured, in that they are not rejected by the tests of the corresponding overidentifying restrictions and the difference Hansen test.

Local or selective separability

Divide plots that correspond to households in different classes depending upon their participation in labor and tenancy markets and estimate reduced form separability equations on each class while controlling for selectivity bias.

Estimating marginal productivity

Estimate a translog production function and test the restrictions on the equality of marginal productivity explicitly

Latent variable approaches

Allow data to sort observations into separable and non-separable classes

Concluding remarks

This article has shown that the rejection of the null-hypothesis of complete markets in household models, based on the widely-used test of the exclusion restrictions implied by separation, can be entirely due to the bias stemming from uncontrolled-for unobservable individual heterogeneity. Our results are particularly important for

plot-level panel datasets where no time dimension is present, since there is *no* means at all, apart from the HT estimator, of testing for separation while controlling for unobservable individual effects (i) if the latter are correlated with the household-level variable that is the focus of the test, and (ii) if no exogenous instruments are available. As was the case with the dataset considered in our empirical illustration, both of these conditions are likely to hold in practice.

The implications of our results are, moreover, suggestive, in that there may be other received results in applied microeconomics, based on panel data, to which the HT estimator could be fruitfully applied. An obvious example is constituted by tests of the precautionary savings motive, in which empirical measures of the risks faced by households are usually time invariant, and in which no attempt is made to correct for unobservable individual effects.

Our results bring the methodology of testing for separation using panel data into sharper focus. This is because we do not reject the null hypothesis of complete markets, *conditional* on λ_{ht} . If one estimates a labor demand function on US individual *firm* data, as in Griliches and Hausman (1986), one finds correlated individual firms effects, as we have found here for households. Thus, by analogy, profit-maximizing behavior by firms is not incompatible with correlated individual effects. However, in our dataset, since labor demand is a function λ_{ht} , it is not independent of household characteristics *per se*, although they are *unobservable* characteristics. Another way of putting this is that, in most panel datasets, testing for separation will undoubtedly uncover correlated individual effects. If separation is taken in its strictest sense to mean that factor demands should be independent of household characteristics, *unconditional* on λ_{ht} , then we do in fact reject the null-hypothesis of complete markets. The key point here revolves around what *type* of household characteristics fall under the λ_h and λ_{ht} headings. If they are made up of household characteristics that only affect labor demand through their impact on the production technology, and the null

of separation is not rejected, then the concept of conditional separation has meaningful operational content. If the first condition is not satisfied, testing for separation in agricultural household models becomes largely devoid of meaning.

A final point concerns the use to which tests of the separation hypothesis are put. Any structural interpretation, in terms of which market failures are binding, of the *pattern* of violations of separation based on observable household characteristics (and thus on those elements of β which are statistically different from zero) will probably be biased unless unobservable individual effects are controlled for using the Hausman-Taylor estimator.

Notes

¹ λ_h represents, for example, time-invariant unobservable household characteristics linked to productivity, whereas λ_{ht} corresponds to transitory unobserved household-specific shocks, that affect both household composition and the average level of human capital in the household. Our null-hypothesis of separability implies that we are formally assuming that λ_h and λ_{ht} are household characteristics that could legitimately affect the production technology and therefore labor demand. More on this issue below, in the conclusion.

²There is a corresponding matrix expression when Z_{ht} involves several household characteristics.

³For simplicity of exposition, we express the instrument set as if the data were balanced. In the empirical application, the unbalanced nature of the data will, of course, be taken into account.

⁴Note that the three-dimensional nature of our dataset allows us to use other combinations of our exogeneous explanatory variables. They are not considered here.

⁵The soil types are clay, red, sandy and barren, with mixed soil types being the excluded category; the crop dummies are other cereals, potatoes, onions, garden vegetables, tomatoes, beetroots, melon and fodder; the excluded category is wheat. We also include a year dummy.

⁶The output and cost shares both equal 1 on plots cultivated by owner-operators. Values strictly less than or greater than one of the ratio obtain on plots under share tenancy contracts.

⁷Note that there are *no* Z_{1ht} variables in this specification.

⁸A household-specific random effects specification (λ_h , not presented) is strongly rejected by the corresponding Hausman test.

⁹We also allowed the covariance matrix of the disturbance term to have household

cluster effects. Such a covariance matrix has the advantage of being flexible, by allowing for arbitrary intra-cluster correlation and heteroskedasticity at the household level. We thus relax the assumption that the correlation within each cluster is constant and has a nested form. Obviously, our less restrictive specification may lead to some losses in terms of efficiency.

¹⁰Note that all other point estimates presented in column 4 are fairly close to those obtained using household-specific fixed effects in column 2, except for that associated with the irrigated plot dummy and the seeds cost share.

¹¹The Shea partial R^2 takes the intercorrelations among the instruments into account.

¹²Obviously, testing our subset of overidentifying restrictions is only valid if the \tilde{X}_{iht} variables are uncorrelated with household-time effects. However, as explained above, this assumption seems to be justified from the theoretical standpoint.

References

- Amemya, T. and T. MaCurdy. 1986. "Instrumental-Variable Estimation of an Error-Components Model." *Econometrica* 54(4):869–881.
- Baltagi, B. H. and S. Khanti-Akom. 1990. "On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators." *Journal of Applied Econometrics* 5(4):401–406.
- Baltagi, B. H., S. H. Song, and B. C. Jung. 2001. "The Unbalanced Nested Error Component Regression Model." *Journal of Econometrics* 101(2):357–381.
- Bardhan, P. K. and C. Udry. 1999. *Development Microeconomics*. Oxford, UK: Oxford University Press.
- Barrett, C. B. 1996. "On Price Risk and the Inverse Farm Size-Productivity Relationship." *Journal of Development Economics* 51(2):193–215.

- Benjamin, D. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." *Econometrica* 60(2):287–322.
- Bowden, R. J. and D. A. Turkington. 1984. *Instrumental Variables*. Cambridge, UK: Cambridge University Press.
- Bowlus, A. and T. Sicular. 2003. "Moving Toward Markets? Labor Allocation in Rural China." *Journal of Development Economics* 71(2):561–583.
- Breusch, T., G. Mizon, and P. Schmidt. 1989. "Efficient Estimation Using Panel Data." *Econometrica* 57(3):695–700.
- Carter, M. and P. Olinto. 2003. "Getting Institutions 'Right' for Whom? Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment." *American Journal of Agricultural Economics* 85(1):173–186.
- Carter, M. and Y. Yao. 2002. "Local versus Global Separability in Agricultural Household Models: The Factor Price Equalization Effect of Land Transfer Rights." *American Journal of Agricultural Economics* 84(3):702–715.
- Cornwell, C. and P. Rupert. 1988. "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators." *Journal of Applied Econometrics* 3(2):149–155.
- Deininger, K. and P. Olinto. 2000. "Why Liberalization Alone Has Not Improved Agricultural Productivity in Zambia: The Role of Asset Ownership and Working Capital Constraints." World Bank Policy Research Working Paper No. 2302, The World Bank, Washington DC.
- Eswaran, M. and A. Kotwal. 1986. "Access to Capital and Agrarian Production Organisation." *Economic Journal* 96(382):482–498.
- Feder, G. 1985. "The Relation Between Farm Size and Farm Productivity: The Role of Family Labor, Supervision, and Credit Constraints." *Journal of Development Economics* 18(2-3):297–313.

- Feder, G., L. J. Lau, J. Y. Lin, and X. Luo. 1990. "The Relationship Between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium." *American Journal of Agricultural Economics* 72(4):1151–1157.
- Gardebroeck, C. and O. Lansink. 2003. "Estimating Farm Productivity Differentials Using Panel Data: The Hausman-Taylor Approach." *Journal of Agricultural Economics* 54(3):397–415.
- Gavian, S. and M. Fafchamps. 1996. "Land Tenure and Allocative Efficiency in Niger." *American Journal of Agricultural Economics* 78(2):460–471.
- Griliches, Z. and J. A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31(1):93–118.
- Hall, A. R., G. Rudebusch, and D. Wilcox. 1996. "Judging Instrument Relevance in Instrumental Variables Estimation." *International Economic Review* 37(2):283–298.
- Hausman, J. and W. Taylor. 1981. "Panel Data and Unobservable Individual Effects." *Econometrica* 49(6):1377–1398.
- Hayashi, F. 2000. *Econometrics*. 1st ed. Princeton NJ: Princeton University Press.
- Hsiao, C. 1986. *Analysis of Panel Data*. Econometric Society Monographs No. 11, 1st ed. Cambridge, UK: Cambridge University Press.
- Jacoby, H. 1993. "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra." *Review of Economic Studies* 60(4):903–921.
- Kevane, M. 1996. "Agrarian Structure and Agricultural Practices in Western Sudan." *American Journal of Agricultural Economics* 78(1):236–245.
- Laffont, J.-J. and M. Matoussi. 1995. "Moral Hazard, Financial Constraints and Sharecropping in El Oulja." *Review of Economic Studies* 62(3):381–399.
- Lambert, S. and T. Magnac. 1994. "Measurement of Implicit Prices of Family Labour in Agriculture: An Application to Côte d'Ivoire." In Caillavet, Gyomard, and

- Lifran, eds., *Agricultural Household Modeling and Family Economics*. Amsterdam, The Netherlands: Elsevier Science.
- Lopez, R. 1984. "Estimating Labour Supply and Production Decisions of Self-Employed Farm Producers." *European Economic Review* 24(1):61–82.
- Matoussi, M. and J. Nugent. 1989. "The Switch to Sharecropping in Medjez-El-Bab." In M. K. Nabli and J. Nugent, eds., *The New Institutional Economics and Development: Theory and Applications to Tunisia*. New York NY: Elsevier Science.
- Sadoulet, E., A. DeJanvry, and C. Benjamin. 1998. "Household Behavior with Imperfect Labor Market." *Industrial Relations* 37(1):85–108.
- Singh, I., L. Squire, and J. Strauss, eds. 1986. *Agricultural Household Models: Extensions, Applications, and Policy*. Baltimore MD: Johns Hopkins University Press.
- Skoufias, E. 1994. "Using Shadow Wages to Estimate Labor Supply of Agricultural Households." *American Journal of Agricultural Economics* 76(2):215–227.
- Sonoda, T. and Y. Maruyama. 1999. "Effects of the Internal Wage on Output Supply: A Structural Estimation for Japanese Rice Farmers." *American Journal of Agricultural Economics* 81(1):131–143.
- Udry, C. 1996a. "Efficiency and Market Structure: Testing for Profit Maximization in African Agriculture." Working paper, Department of Economics, Northwestern University.
- . 1996b. "Gender, Agricultural Production, and the Theory of the Household." *Journal of Political Economy* 104(5):1010–1046.
- Varian, H. 1978. *Microeconomic Analysis*. 1st ed. New York NY: Wiley.

Table 1: Tests of the Separation Hypothesis Based on Exclusion Restrictions on Household Demographic Characteristics

Authors	Dataset	Exclusion restrictions	Estimated equation(s)	Separability	
				Global	Local
Pitt and Rosenzweig, 1986	Indonesia, 1978/86	Family morbidity (no. of sick days)	Farm profits	Not rejected	Not tested
Deolalikar, 1988	India, 1976/78	Hh. size and weight-for-height	Farm outputs	Rejected	Not tested
Benjamin, 1992	Java, 1980	Hh. size and composition	Hh. labor demand	Not rejected	Not tested
Bowlus and Sicular, 2003	China, 1990-1993	Hh. size and composition	Hh. labor demand	Rejected	Not tested
Grimard, 2000	Côte d'Ivoire, 19XX	Hh. composition	Farm labor demand	Rejected	
Lopez, 2004	Canada, 1970 (agg. data)			Rejected	

Table 2: Tests of the Separation Hypothesis Based on Exclusion Restrictions on Household Demographic and Other Characteristics.

Authors	Dataset	Exclusion restrictions	Estimated equation(s)	Separability	
				Global	Local
Benjamin, 1994	Java, 1980	Area harvested (previous year)	Farm output, labor demand		
Carter and Yao, 2002	China, 1988, 1993	Land-labor endowment ratio	Labor demand	Rejected	Not tested
Carter and Olinto, 2003	Paraguay, 1991, 94	Land property rights	Cap. invest. demand	Not tested	Not tested
Feder et al, 1990	China, 1987	Liquid assets, hh. size, size of hh. labor force	Output supply	Not tested	Not tested
Udry, 1996	Burkina Faso, 1981-85, Kenya, 1985-87	Hh. size, non-farm wealth, total area on other plots	Plot output, plot labor demand	Rejected	
Sadoulet et al, 1998	Mexico, 1994	Hh. endowment of unskilled and skilled labor, hh. migration assets, assets affecting utility	Labor intensity	Not tested	Not tested
Kevane, 1996	Sudan, 1990	Wealth, land and labor endowments	Yield per hectare		
Gavian and Fafchamps, 1996	Niger, 1990/91	Hh. manpower, proxies for hh. wealth	Qty. manure on field	Rejected	
Vakis et al, 2004	Peru, 1997	Hh labor endowment and cons. chars., hours worked off-farm	Hh.'s on-farm work	Rejected	Not rejected

Table 3: Summary Statistics, ElOulja, Tunisia (447 Plots (i), 150 Households (h), 196 Household-Years (ht))

	Mean	Median	Std. dev.
Person-day labor input per hectare (Y_{iht})	190.860	119.0	253.271
Plot characteristics (X_{1iht})			
Soil type 1 (clay)	0.190	0.0	0.393
Soil type 2 (red)	0.201	0.0	0.401
Soil type 3 (sandy)	0.446	0.0	0.497
Soil type 4 (barren)	0.058	0.0	0.235
Irrigated plot	0.882	1.0	0.322
Contractual terms (X_{2iht})			
<div style="display: flex; align-items: center;"> <div style="flex: 1;"> <div>% of costs paid by the cultivator</div> <div>% of output accruing to the cultivator</div> </div> <div style="flex: 1; text-align: center;">for:</div> </div>			
Manure	1.008	1.0	0.121
Chemical fertilizer	1.016	1.0	0.145
Irrigation	0.999	1.0	0.150
Plowing	0.984	1.0	0.250
Family labor	1.063	1.0	0.243
Hired labor	1.041	1.0	0.230
Seeds	1.008	1.0	0.094
Transportation	1.006	1.0	0.171
Surface of plot in hectares	5.615	1.5	13.535
Household size (Z_{2ht})	8.257	7.0	5.117

Table 4: Labor Demand Equations: Pooling, Fixed Effects, Random Effects, and Hausman-Taylor Estimators (447 Plots (i), 150 Households (h), 196 Household-Years (ht))

Mean of dep. var. = 3.825	Pooling	Fixed effects λ_h	Random effects λ_{ht}	HT (efficient) λ_{ht}
	(1)	(2)	(3)	(4)
Household characteristics (Z_{2ht})				
Log household size	0.571 (4.55)	0.943 (2.56)	0.516 (3.12)	0.404 (0.71)
	()	()	()	()
	()	()	()	()
	()	()	()	()
	()	()	()	()
	()	()	()	()
Plot characteristics (X_{1iht})				
Soil type 1 (clay)	-0.013 (-0.04)	0.206 (0.66)	0.184 (0.60)	0.248 (0.60)
Soil type 2 (red)	-0.467 (-1.40)	-0.304 (-1.07)	-0.471 (-1.60)	-0.510 (-0.94)
Soil type 3 (sandy)	-0.171 (-0.62)	-0.081 (-0.30)	-0.057 (-0.20)	-0.006 (0.02)
Soil type 4 (barren)	0.305 (0.84)	0.546 (1.48)	0.179 (0.46)	0.633 (1.2)
Irrigated plot	0.577 (1.92)	0.415 (1.78)	0.412 (1.86)	1.017 (2.49)
Contractual terms (X_{2iht})				
% of costs paid by the cultivator for: % of output accruing to the cultivator				
Manure	0.975 (1.31)	4.120 (4.33)	1.346 (1.48)	1.295 (1.14)
Chemical fertilizer	1.070 (1.49)	-0.685 (-0.79)	1.191 (1.39)	1.979 (2.34)
Irrigation	-0.128 (-0.25)	1.826 (2.74)	0.233 (0.35)	0.539 (1.53)
Plowing	-0.080 (-0.26)	-1.542 (-3.31)	-0.408 (-1.07)	-0.864 (-1.64)
Family labor	0.776 (2.38)	-0.080 (-0.15)	0.645 (1.36)	0.170 (0.38)
Hired labor	-0.620 (-2.07)	-0.417 (-0.79)	-0.734 (-1.48)	-0.637 (-1.84)
Seeds	-0.675 (-0.83)	-2.902 (-2.71)	-0.758 (-0.81)	0.279 (0.21)
Transportation	-1.074 (-1.94)	-0.981 (-1.71)	-1.403 (-2.58)	-1.494 (-2.71)
Log surface of plot in hectares	-1.008 (-11.10)	-0.474 (-5.70)	-0.816 (-11.59)	-0.681 (-6.20)
Crop choice hazard rate	()	()	()	()
Joint signif. of cost shares: F -statistic [p -value]	4.96 [0.001]	4.20 [0.000]	16.00 [0.042]	29.15 [0.000]
R^2	0.6934	0.6558	0.6856	$n.a.$
Test of overid. restrictions [d.f., p -value]	$n.a.$	$n.a.$	$n.a.$	10.805 [12,0.545]
Shea Partial R^2 for log household size	$n.a.$	$n.a.$	$n.a.$	0.071
Canonical correlation	$n.a.$	$n.a.$	$n.a.$	0.053 [0.002]
Difference Hansen test [d.f., p -value]	$n.a.$	$n.a.$	$n.a.$	10.214 [13,0.676]